2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Exploiting Convexity in Direct Optimal Control: a Sequential Convex Quadratic Programming Method

Robin Verschueren[1], Niels van Duijkeren[2], Rien Quirynen[1,3], Moritz Diehl[1]

***Abstract*— Direct optimal control methods first discretize a continuous-time Optimal Control Problem (OCP) and then solve the resulting Nonlinear Program (NLP). Sequential Quadratic Programming (SQP) is a popular family of algorithms to solve this finite dimensional optimization problem. In the specific case of a least squares cost, the Generalized Gauss-Newton (GGN) method is a popular approach which works very well under some assumptions. This paper proposes a Sequential Convex Quadratic Programming (SCQP) scheme which exploits additional convexities in the NLP in order to generalize the GGN algorithm, possibly extend its applicability and improve its local convergence. These properties are studied in detail for the proposed SCQP algorithm, which will be compared to the classical GGN method using a numerical case study of the optimal control of an inverted pendulum.**

I. INTRODUCTION

We are interested in solving Nonlinear Programs (NLP) of the following form, which typically arise from the multiple-shooting discretization [5] of a continuous-time Optimal Control Problem (OCP):

$$\min_{\substack{x_0,\dots,x_N\\u_0,\dots,u_{N-1}}} \ \frac{1}{2}\sum_{k=0}^{N-1}\|h_k(x_k,u_k)\|_2^2 \ + \ \frac{1}{2}\|h_N(x_N)\|_2^2 \tag{1a}$$

$$\text{s.t.} \quad x_0 = \overline{x}_0, \tag{1b}$$

$$x_{k+1} = f(x_k,u_k), \quad k = 0,\dots,N-1, \tag{1c}$$

$$e_k(x_k,u_k) \leq \mathbb{0}, \qquad k = 0,\dots,N-1, \tag{1d}$$

$$e_N(x_N) \leq \mathbb{0}, \tag{1e}$$

with optimization variables $x_k \in \mathbb{R}^{n_x}, u_k \in \mathbb{R}^{n_u}$ and a least squares objective defined by $h_k : \mathbb{R}^{n_x}\times\mathbb{R}^{n_u} \to \mathbb{R}^{n_h}$, $h_N : \mathbb{R}^{n_x} \to \mathbb{R}^{n_h}$. In addition, $f : \mathbb{R}^{n_x}\times\mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ is the state transition map which is typically obtained by numerical integration. Functions $e_k : \mathbb{R}^{n_x}\times\mathbb{R}^{n_u} \to \mathbb{R}^{n_c}$, $e_N : \mathbb{R}^{n_x} \to \mathbb{R}^{n_c}$ define the inequality constraints and $\overline{x}_0 \in \mathbb{R}^{n_x}$ denotes the initial value for the state vector. In case of receding horizon real-time control [13], an NLP of this form (1) is solved at each sampling instant where $\overline{x}_0$ denotes the current state estimate.

A popular technique to solve NLPs with a least-squares objective, such as Eq. (1a), is the Generalized Gauss-Newton (GGN) method proposed by [4]. This Sequential Quadratic Programming (SQP) method uses a Gauss-Newton Hessian approximation for each Quadratic Programming (QP) subproblem. Advantages of the GGN method include that no second order derivatives need to be evaluated, that it is a multiplier-free algorithm, and that the Hessian approximation is always positive (semi-) definite. Therefore each QP subproblem is convex and can be solved efficiently and reliably.

[1]Systems Control and Optimization Laboratory, Department of Microsystems Engineering, University of Freiburg, Germany robin.verschueren@imtek.uni-freiburg.de
[2]Division of Production engineering, Machine design and Automation, Department of Mechanical Engineering, KU Leuven, Belgium
[3]Department of Electrical Engineering (ESAT), KU Leuven, 3001 Heverlee, Belgium.

The GGN algorithm has been shown to be a reliable approach for real-time Nonlinear Model Predictive Control (NMPC) [8].

It is known that the GGN algorithm has good local convergence properties when the residual functions $h_k(x_k^\star,u_k^\star)$ are small at the solution [7]. Whenever this is not the case, the second order derivatives for either the residual or for the constraint functions in (1c)-(1e) need to be evaluated and included in the Hessian to obtain convergence even when initializing the algorithm arbitrarily close to a local minimizer. In case of an exact Hessian based method, one always obtains contraction close to a minimizer and this convergence speed is quadratic [17]. However, the Hessian of the Lagrangian can be indefinite such that the corresponding QP subproblem is non-convex and therefore generally not easy to solve. Many different remedies have been proposed, such as Hessian regularization techniques [15], [17] or the addition of an equality constrained QP phase [14].

Unlike the latter approaches which are based on the approximation of an indefinite QP, we aim at directly formulating convex subproblems. Similar to the family of Sequential Convex Programming (SCP) methods as discussed in [19], one can often exploit some convexity in either the objective or the constraint functions. The idea of an SCP method is to linearize all non-convex functions and solve the resulting convex subproblem in each iteration. Even though this approach can indeed locally result in good linear convergence [18], one needs to rely on a general convex solver. In this paper, we instead propose an SQP method in which we use the convexity available from objective and constraint functions to obtain a more accurate Hessian approximation. This Hessian is based on the second order derivatives of convex functions and is therefore always positive semidefinite, similar to the case for the GGN method. A similar method for unconstrained problems is presented in [12].

The contribution of this paper concerns the proposed Sequential Convex Quadratic Programming (SCQP) scheme, which is presented as a generalization of the classical GGN method. The advantages of this algorithm are motivated from the computational point of view, since the second order derivatives needed for each Hessian approximation are relatively easy to evaluate [11] unlike the propagation of second order derivatives for the dynamics (1c). In addition, each subproblem is a convex QP which is typically easier to solve than a more general convex subproblem as done by SCP methods. Note that because of these reasons, SCQP is suited for real-time optimization, e.g. in an NMPC setting.

Furthermore, it will be shown that SCQP can indeed improve the local convergence properties as compared to the GGN method, possibly resulting in a stronger contraction rate. The performance of this scheme is illustrated using the numerical case study considering an optimal control problem for an inverted pendulum.

II. PROBLEM FORMULATION

In order to present our method in a more compact way, we introduce a slightly more general NLP formulation instead of the

NLP from Eq. (1). It should be noted that a non-convex inequality constraint can always be rewritten as $\phi_i(c_i(w)) \leq 0$, with ϕ_i convex, such that the optimization problem reads as:

$$\min_{w\in\mathbb{R}^n} \quad \phi_0(c_0(w)) \tag{2a}$$
$$\text{s.t.} \quad g_i(w) = 0, \qquad i = 1,\ldots,p, \tag{2b}$$
$$\phi_i(c_i(w)) \leq 0, \, i = 1,\ldots,q, \tag{2c}$$

with convex output functions $\phi_i : \mathbb{R}^m \to \mathbb{R}$ and possibly nonlinear functions $c_i : \mathbb{R}^n \to \mathbb{R}^m$ for $i = 0,\ldots,q$ (for simplicity of notation, we assume the output dimension of all c_i to be equal to m). GGN is well suited for the case $\phi_0(c_0(w)) = \frac{1}{2}\|c_0(w)\|_2^2$ as it exploits the least-squares nature of the objective function. SCQP extends this idea to exploit general convex output functions in the objective as well as in the inequality constraints.

To arrive at an even more compact notation for NLP (2), we will refer to the functions $\phi_i(c_i(w))$ as $\psi_i(w)$. These functions are generally non-convex and further assumed to be three times continuously differentiable.

We define the Lagrangian of NLP (2) as

$$\mathcal{L}(w,\lambda,\mu) := \psi_0(w) + \lambda^\top g(w) + \mu^\top \psi(w), \tag{3}$$

with the Lagrange multipliers $\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q$, and $g : \mathbb{R}^n \to \mathbb{R}^p, \psi : \mathbb{R}^n \to \mathbb{R}^q$ respectively denote the concatenation of equality and inequality constraints. Furthermore, we define $\gamma(w) := [g(w)^\top, \psi_{i\in\mathcal{A}}(w)^\top]^\top$, as the vector of equality and active inequality constraints, with $\mathcal{A}$ the set of n_a inequality constraints which are active in w. The Lagrange multipliers corresponding to constraints $\gamma(w)$ are denoted by $\nu = [\lambda^\top, \mu_{i\in\mathcal{A}}^\top]^\top \in \mathbb{R}^{p+n_a}$.

When the NLP (2) is solved with a full-step SQP method, the solution iterates evolve according to

$$\begin{bmatrix} w_{k+1} \\ \lambda_{k+1} \\ \mu_{k+1} \end{bmatrix} = \begin{bmatrix} w_k \\ \lambda_k \\ \mu_k \end{bmatrix} + \begin{bmatrix} s_k \\ s_k^\lambda \\ s_k^\mu \end{bmatrix}. \tag{4}$$

The step s_k and the multipliers λ_{k+1}, μ_{k+1} are the primal and dual solution, respectively, of the following QP subproblem

$$\min_{s_k\in\mathbb{R}^n} \quad \frac{1}{2} s_k^\top B_k\, s_k + \nabla\psi_0(w_k)^\top s_k \tag{5a}$$
$$\text{s.t.} \quad \nabla g_i(w_k)^\top s_k + g_i(w_k) = 0, \quad i = 1,\ldots,p, \tag{5b}$$
$$\nabla \psi_i(w_k)^\top s_k + \psi_i(w_k) \leq 0, \quad i = 1,\ldots,q. \tag{5c}$$

For an exact Newton scheme, we use the exact Hessian matrix $B_k = \nabla_w^2 \mathcal{L}(w_k, \lambda_k, \mu_k)$. As stated before, in the case of a least squares cost function $\psi_0(w) = \frac{1}{2}\|c_0(w)\|_2^2$ one often uses the Gauss-Newton Hessian approximation [4] instead:

$$B_k^{\text{GN}} = J(w_k)^\top J(w_k), \text{ where } J(w_k) = \frac{\partial c_0(w_k)}{\partial w}. \tag{6}$$

A convenient property is that $B_k^{\text{GN}} \succeq \mathbb{0}$. However, a local minimizer $(w^\star, \lambda^\star, \mu^\star)$ of NLP (2) might become an unstable point for the SQP iterations with a Gauss-Newton Hessian approximation, i.e. the SQP method in some cases may not converge to $(w^\star, \lambda^\star, \mu^\star)$ even when initialized arbitrarily close to this minimizer [6], [9].

Motivated by this observation, the next section presents a necessary and sufficient condition on the Hessian approximation for asymptotic stability of a local minimizer $w^\star$.

III. Asymptotic Stability of a Local Minimizer

First, we consider the case of unconstrained optimization. Afterwards, we show that the same result is applicable to constrained optimization problems. We assume that the Hessian approximation $B(w_k)$ is continuously differentiable in w_k. Furthermore, unless it is specified, we will assume that $B(w_k) \succ \mathbb{0}$.

A. Unconstrained Optimization Problem

Consider the simplified version of the NLP in (2) without any constraints:

$$\min_{w\in\mathbb{R}^n} \quad \psi_0(w). \tag{7}$$

Assume that the second order sufficient conditions (SOSC) for optimality hold at a local minimizer $w^\star$, i.e. $\nabla^2\psi_0(w^\star) \succ 0$. We solve the first order necessary optimality condition $\nabla\psi_0(w) = 0$ with a Newton-type SQP method, resulting in iterations

$$\begin{aligned} w_{k+1} &= F(w_k) \\ &= w_k + \arg\min_{s_k\in\mathbb{R}^n} \frac{1}{2} s_k^\top B(w_k) s_k + \nabla\psi_0(w_k)^\top s_k \\ &= w_k - B(w_k)^{-1}\nabla\psi_0(w_k). \end{aligned} \tag{8}$$

A standard result from linear stability analysis is stated without proof in the following lemma.

Lemma 1 (Linear Stability Analysis): Regard an iteration of the form $w_{k+1} = F(w_k)$ with F a continuously differentiable function in a neighborhood of a fixed point $F(w^*) = w^*$. If all eigenvalues of the Jacobian $\frac{\partial F}{\partial w}(w^*)$ have a modulus smaller than one, i.e. if the spectral radius is smaller than one $\rho\left(\frac{\partial F}{\partial w}(w^*)\right) < 1$, then the fixed point w^* is asymptotically stable. In that case, when started in a neighborhood of the fixed point, the iterates converge to w^* with a Q-linear convergence rate with asymptotic contraction rate $\rho\left(\frac{\partial F}{\partial w}(w^*)\right)$. On the other hand, if one of the eigenvalues has a modulus larger than one, i.e. if $\rho\left(\frac{\partial F}{\partial w}(w^*)\right) > 1$, then the fixed point is unstable.

The Taylor expansion of $F(w_k)$ in (8) around $w^\star$ reads as

$$\begin{aligned} w_{k+1} = w_k - &B(w^\star)^{-1}\nabla^2\psi_0(w^\star)(w_k - w^\star) \\ &+ \mathcal{O}(\|w_k - w^\star\|^2), \end{aligned} \tag{9}$$

where we used the fact that $\nabla\psi_0(w^\star) = \mathbb{0}$.

Neglecting higher order terms, we can rewrite (9) as

$$\Delta w_{k+1} = \underbrace{B(w^\star)^{-1}(B(w^\star) - \nabla^2\psi_0(w^\star))}_{M^\star} \Delta w_k, \tag{10}$$

with $\Delta w_k = w_k - w^\star$.

Note that from (10) it follows that $\frac{\partial F}{\partial w}(w^\star) = M^\star$. A different characterization of the necessary and sufficient condition on the spectral radius in Lemma 1 is stated in the following lemma.

Lemma 2: Define $M^\star$ as in (10). Then the two following statements are equivalent.

1) The spectral radius $\rho(M^\star) \leq \alpha$,
2) $-\alpha B(w^\star) \preceq B(w^\star) - \nabla^2\psi_0(w^\star) \preceq \alpha B(w^\star)$.

Proof: By assumption, $B(w^\star) \succ \mathbb{0}$, so $B(w^\star)^{-\frac{1}{2}}$ exists. It follows that the eigenvalues of $M^\star$ and $\Sigma := B(w^\star)^{-\frac{1}{2}}(B(w^\star) - \nabla^2\psi_0(w^\star))B(w^\star)^{-\frac{1}{2}}$ are the same. Assume now that $\rho(M^\star) \leq \alpha$. As Σ is symmetric, we can write that $-\alpha\mathbb{1} \preceq \Sigma \preceq \alpha\mathbb{1}$ and thus $-\alpha B(w^\star) \preceq B(w^\star) - \nabla^2\psi_0(w^\star) \preceq \alpha B(w^\star)$. The converse follows from the definition of spectral radius. ■

Motivated by Lemmas 1 and 2, we can now state a necessary and sufficient condition on $B(w^\star)$ in order for $w^\star$ to be asymptotically stable with contraction rate α.

Theorem 3: A solution $w^\star$ of NLP (7) is an asymptotically stable fixed point for the Newton-type iteration in (8) with asymptotic

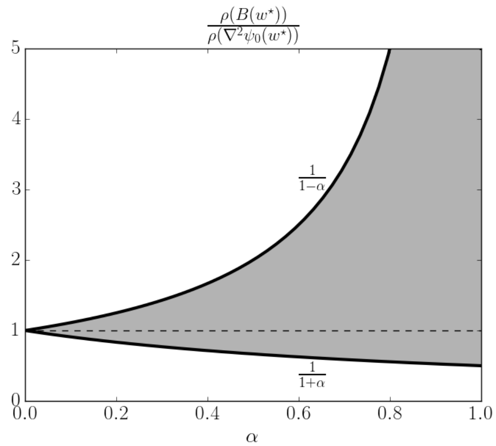


Fig. 1: For a solution $w^\star$ to be stable, the Hessian approximation $B(w^\star)$ has to lie inside the shaded region (11).

contraction rate $0 \le \alpha < 1$, if and only if the following conditions hold:

$$B(w^\star) \succeq \frac{\nabla^2\psi_0(w^\star)}{1+\alpha}, \tag{11a}$$

$$B(w^\star) \preceq \frac{\nabla^2\psi_0(w^\star)}{1-\alpha}. \tag{11b}$$

Proof: These two conditions are equivalent to

$$-\alpha B(w^\star) \preceq B(w^\star) - \nabla^2\psi_0(w^\star) \preceq \alpha B(w^\star). \tag{12}$$

This condition is equivalent to $\rho(\frac{\partial F}{\partial w}(w^\star)) \le \alpha < 1$ because of Lemma 2, which is in turn the necessary and sufficient condition for asymptotic stability as defined by Lemma 1. ■

The bounds in (11) are illustrated in Fig. 1.

B. Equality and Inequality constraints

Let us return to the Newton-type SQP method from Eq. (5), applied to the original NLP (2). Assume that the linear independence constraint qualification (LICQ), SOSC and strict complementarity hold at the local minimizer $(w^\star, \lambda^\star, \mu^\star)$ as defined in [16]. It follows that the active set for the QP solution is stable close to a local minimizer of the NLP, i.e. the active set for the QP in (5) is also the active set of the original NLP [9].

The KKT system corresponding to the QP subproblem (5) after fixing the active inequality constraints and omitting the inactive ones, reads as

$$\begin{bmatrix} B(w,\nu) & G(w)^\top \\ G(w) & \mathbb{0} \end{bmatrix} \begin{bmatrix} s \\ s^\nu \end{bmatrix} = -\begin{bmatrix} \nabla_w \mathcal{L}(w,\nu) \\ \gamma(w) \end{bmatrix}, \tag{13}$$

where $G(w) = \frac{\partial \gamma}{\partial w}(w)$ and we recall that the multipliers ν correspond to the active constraints $\gamma(w) := [g(w)^\top, \psi_{i\in\mathcal{A}}(w)^\top]^\top$ only.

Let us denote the constraint matrix at the local solution by $G^\top := G(w^\star)^\top$, such that its QR factorization reads:

$$G^\top = Q\overline{R} = \begin{bmatrix} Y & Z \end{bmatrix} \begin{bmatrix} R \\ \mathbb{0} \end{bmatrix}, \tag{14}$$

with orthogonal matrix $\begin{bmatrix} Y & Z \end{bmatrix}$. It follows that the matrix Z is a null space for the constraint matrix G at the solution, i.e. $GZ = \mathbb{0}$. Let us define the reduced Hessian matrix and its corresponding approximation at a local solution $(w^\star, \nu^\star)$:

$$\Lambda^\star := Z^\top \nabla_w^2 \mathcal{L}^\star Z,$$
$$\widetilde{B}^\star := Z^\top B^\star Z,$$

with shorthands $\nabla_w^2\mathcal{L}^\star := \nabla_w^2\mathcal{L}(w^\star,\nu^\star)$ and $B^\star := B(w^\star,\nu^\star)$. By assumption, SOSC holds at the local minimizer $(w^\star,\nu^\star)$, which implies that the reduced Hessian $\Lambda^\star \succ \mathbb{0}$.

Introducing the compact notation $z = [w^\top, \nu^\top]^\top$, one step of the Newton-type SQP method (13) reads as

$$\begin{aligned} z_{k+1} &= z_k - \begin{bmatrix} B(z_k) & G(w_k)^\top \\ G(w_k) & \mathbb{0} \end{bmatrix}^{-1} \begin{bmatrix} \nabla_w \mathcal{L}(z_k) \\ \gamma(w_k) \end{bmatrix} \\ &= z_k - \begin{bmatrix} B^\star & G^\top \\ G & \mathbb{0} \end{bmatrix}^{-1} \begin{bmatrix} \nabla_w^2 \mathcal{L}^\star & G^\top \\ G & \mathbb{0} \end{bmatrix} (z_k - z^\star), \end{aligned} \tag{15}$$

after neglecting higher order terms. Here, we used a Taylor expansion of $\begin{bmatrix} \nabla_w \mathcal{L}(z_k) \\ \gamma(w_k) \end{bmatrix}$ around $z^\star$ and the fact that this quantity vanishes at $z^\star$. We can now state a version of Theorem 3, including equality and inequality constraints.

Theorem 4: Assume we are close to a local minimizer $(w^\star, \lambda^\star, \mu^\star)$ of NLP (2) where LICQ, SOSC and strict complementarity hold, such that the current active set is equal to the active set at the local minimizer. Then $(w^\star, \nu^\star)$ is asymptotically stable for the Newton-type iteration (15) with asymptotic contraction rate $0 \le \alpha < 1$, if and only if the following conditions hold:

$$\widetilde{B}^\star \succeq \frac{\Lambda^\star}{1+\alpha}, \tag{16a}$$

$$\widetilde{B}^\star \preceq \frac{\Lambda^\star}{1-\alpha}. \tag{16b}$$

Proof: Let us define a similar change of variables as in (10), $\Delta\widetilde{z}_k = U^\top(z_k - z^\star)$, with orthogonal matrix

$$U = \begin{bmatrix} Z & Y & \\ & & \mathbb{1} \end{bmatrix}, \tag{17}$$

where Z, Y are defined by the QR factorization in (14). Equation (15) then reads as:

$$\begin{aligned} \Delta\widetilde{z}_{k+1} &= U^\top \begin{bmatrix} B^\star & G^\top \\ G & \mathbb{0} \end{bmatrix}^{-1} UU^\top \begin{bmatrix} B^\star - \nabla_w^2\mathcal{L}^\star & 0 \\ 0 & 0 \end{bmatrix} U \Delta\widetilde{z}_k, \\ &= L^{-1} P\, \Delta\widetilde{z}_k, \end{aligned}$$

where we defined the matrices

$$L = U^\top \begin{bmatrix} B^\star & G^\top \\ G & \mathbb{0} \end{bmatrix} U,$$
$$P = U^\top \begin{bmatrix} B^\star - \nabla_w^2\mathcal{L}^\star & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{bmatrix} U.$$

For each eigenvalue β of matrix $L^{-1}P$ there exists a $v \ne \mathbb{0}$ satisfying $Pv = \beta Lv$. Expanding the matrix products yields

$$\begin{aligned} &\beta \begin{bmatrix} Z^\top B^\star Z & Z^\top B^\star Y & \mathbb{0} \\ Y^\top B^\star Z & Y^\top B^\star Y & R \\ \mathbb{0} & R^\top & \mathbb{0} \end{bmatrix} \begin{bmatrix} v_z \\ v_y \\ v_r \end{bmatrix} = \\ &\begin{bmatrix} Z^\top(B^\star - \nabla_w^2\mathcal{L}^\star)Z & Z^\top(B^\star - \nabla_w^2\mathcal{L}^\star)Y & \mathbb{0} \\ Y^\top(B^\star - \nabla_w^2\mathcal{L}^\star)Z & Y^\top(B^\star - \nabla_w^2\mathcal{L}^\star)Y & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} \end{bmatrix} \begin{bmatrix} v_z \\ v_y \\ v_r \end{bmatrix}, \end{aligned} \tag{18}$$

where $GZ = \mathbb{0}$ and $GY = R^\top$ have been used.

For $\beta \ne 0$, from the bottom row of the equation in (18) we have that $\beta R^\top v_y = \mathbb{0}$, which implies $v_y = 0$ since R is invertible.

From the top row it then follows that $\beta Z^\top B^\star Z v_z = Z^\top (B^\star - \nabla_w^2 \mathcal{L}^\star) Z v_z$, and consequently that

$$\beta v_z = (\underbrace{Z^\top B^\star Z}_{\widetilde{B}^\star})^{-1}(\underbrace{Z^\top B^\star Z}_{\widetilde{B}^\star} - \underbrace{Z^\top \nabla_w^2 \mathcal{L}^\star Z}_{\Lambda^\star}) v_z. \tag{19}$$

Thus, the nonzero eigenvalues of matrix $L^{-1}P$ are equal to the eigenvalues of $(\widetilde{B}^\star)^{-1}(\widetilde{B}^\star - \Lambda^\star)$. We have now recovered the same form of (10) for each Newton-type iteration. Consequently, the rest of the proof follows that of Theorem 3 for the unconstrained case. ■

As a corollary of Theorem 4 it holds that $\widetilde{B}^\star \succ \frac{1}{2}\Lambda^\star$ implies asymptotic stability for the local minimizer $z^\star$. This can be seen from taking the limit of (16a) for $\alpha \to 1$ (see Fig. 1). Motivated by these results, we introduce a novel Hessian approximation in the next section.

IV. Sequential Convex Quadratic Programming

Consider the NLP (2), which is the more general form of OCP (1), and we recall that ϕ_i for $i = 0, \dots, q$ are convex. We construct a Hessian approximation using the contributions from the functions we know to be convex. This is motivated by the fact that SQP might yield large steps s_k due to the linearization of the inequalities, whose convexity is ignored. We propose to use a modification of the Generalized Gauss-Newton method, where we linearize the convex inequalities, but add their positive definite second order derivative to the Hessian. This Hessian approximation then reads

$$\begin{aligned} B^{\mathrm{SCQP}}(w,\mu) := {}& \frac{\partial c_0}{\partial w}(w)^\top \nabla_c^2 \phi_0(c_0(w)) \frac{\partial c_0}{\partial w}(w) \\ &+ \sum_{i=1}^{q} \mu_i \frac{\partial c_i}{\partial w}(w)^\top \nabla_c^2 \phi_i(c_i(w)) \frac{\partial c_i}{\partial w}(w). \end{aligned} \tag{20}$$

We will refer to the SQP method using B^{SCQP} as a Hessian approximation in its QP (5) subproblem, as *Sequential Convex Quadratic Programming* (SCQP). The corresponding Hessian approximation error reads as:

$$\begin{aligned} E^{\mathrm{SCQP}}(w,\lambda,\mu) = {}& \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(w) + \sum_{j=1}^{m} \frac{\partial \phi_0}{\partial c_{0,j}}(w) \nabla^2 c_{0,j}(w) \\ &+ \sum_{i=1}^{q} \mu_i \sum_{j=1}^{m} \frac{\partial \phi_i}{\partial c_{i,j}}(w) \nabla^2 c_{i,j}(w). \end{aligned} \tag{21}$$

Note that the GGN method is a special case of this class of methods. In comparison to the Gauss-Newton Hessian, B^{SCQP} has the benefit that we are closer to a Hessian approximation $\widetilde{B}^\star \succ \frac{1}{2}\Lambda^\star$ that implies asymptotic stability of a local minimizer $w^\star$. This fact is a corollary from the following lemma.

Lemma 5: For a local minimum $(w^\star, \lambda^\star, \mu^\star)$ of NLP (2) with least squares cost function $\phi_0(c_0(w)) = \frac{1}{2}\|c_0(w)\|_2^2$, it holds that

$$B^{\mathrm{SCQP}}(w^\star, \mu^\star) \succeq B^{\mathrm{GN}}(w^\star). \tag{22}$$

Proof: For a least squares cost function, Eq. (22) follows directly from

$$\begin{aligned} B^{\mathrm{SCQP}}(w,\mu) &= \frac{\partial c_0}{\partial w}(w)^\top \frac{\partial c_0}{\partial w}(w) \\ &\quad + \sum_{i=1}^{q} \mu_i \frac{\partial c_i}{\partial w}(w)^\top \nabla_c^2 \phi_i(c_i(w)) \frac{\partial c_i}{\partial w}(w), \\ &= B^{\mathrm{GN}}(w) + \sum_{i=1}^{q} \mu_i \frac{\partial c_i}{\partial w}(w)^\top \nabla_c^2 \phi_i(c_i(w)) \frac{\partial c_i}{\partial w}(w), \end{aligned}$$

and the fact that inequalities ϕ_i are convex and multipliers μ_i are nonnegative at a local solution. ■

Remark 6: SCQP is motivated by the results of Theorem 4. However, we do not guarantee that the SCQP Hessian satisfies the bounds (16) in general. As such, local convergence, just as in the case of GGN, is not guaranteed. In practice, the SCQP Hessian is often 'closer' to the exact Hessian, which might result in better convergence properties, as shown in Section V.

In the case of constant second order derivatives for the objective and inequality constraint functions, SCQP is specifically easy to implement and computationally cheap. But also in general, the second order derivatives for the Hessian in (20) can be efficiently evaluated using Algorithmic Differentiation (AD) [11].

An alternative view on SCQP is the following. Applying Sequential Convex Programming (SCP) in order to solve NLP (2) results in the subproblems:

$$\min_{w \in \mathbb{R}^n} \quad \phi_0\left(\frac{\partial c_0}{\partial w}(w_k)(w - w_k) + c_0(w_k)\right) \tag{23a}$$

$$\text{s.t.} \quad \frac{\partial g_i}{\partial w}(w_k)(w - w_k) + g_i(w_k) = 0, \tag{23b}$$

$$\phi_j\left(\frac{\partial c_j}{\partial w}(w_k)(w - w_k) + c_j(w_k)\right) \le 0, \tag{23c}$$

with $i = 1, \dots, p$, $j = 1, \dots, q$ and the current linearization point w_k. One iteration of the SCQP method is then equivalent to an exact Hessian based SQP iteration for the latter convex subproblem (23).

We illustrate the benefits of the SCQP algorithm using a numerical case study in the next section.

V. Numerical Example

In this section, we solve an OCP of the form in (1) using Sequential Quadratic Programming. More specifically, we compare different techniques of Hessian approximation and their resulting local convergence properties, including the proposed SCQP scheme and the more classical Gauss-Newton (GGN) and exact Hessian (EH) based SQP method.

A. Implementation and software

To numerically solve the optimal control problems, we adopt the open-source CasADi [2] software framework, which has been proven to solve OCPs reliably and efficiently [3]. More specifically, we use the Python front-end to formulate the OCP as a nonlinear program (NLP) using direct multiple shooting [5] as a discretization method.

The resulting NLP is passed to the open-source solver IPOPT [20]. It implements a primal-dual interior point method suited for solving large-scale NLPs. The linear algebra subroutine calls were passed to the sparse solver `ma86` from the HSL library [1]. For SCQP, GGN and exact Hessian SQP, we use the open-source QP solver qpOASES [10]. Note that the software packages mentioned above can be conveniently called from within the CasADi framework.

B. Inverted pendulum with terminal region

As an example, we regard a pendulum on a cart, as in Fig. 2. In the following, we neglect the mass of the rod. The control objective is to steer the pendulum with mass m [kg] inside a circular region, in a control horizon of $1\,\mathrm{s}$. We denote the horizontal position and velocity of the cart with $p\,[\mathrm{m}]$ and $v\,[\mathrm{m/s}]$, and the angle and angular velocity with $\theta\,[\mathrm{rad}]$ and $\omega\,[\mathrm{rad/s}]$, respectively. The

TABLE I: Parameters for the inverted pendulum example

Parameter	Description	Value
M	mass of cart	$1\,\mathrm{kg}$
m	mass of pendulum	$0.1\,\mathrm{kg}$
l	length of rod	$0.8\,\mathrm{m}$
g	gravitational acceleration	$9.81\,\mathrm{kg\,m/s^2}$

dynamics then read as

$$\begin{bmatrix}\dot{p}\\ \dot{v}\\ \dot{\theta}\\ \dot{\omega}\end{bmatrix} = \begin{bmatrix} v \\ \frac{-ml\sin(\theta)\omega^2 + mg\cos(\theta)\sin(\theta)+F}{M+m-m(\cos(\theta))^2} \\ \omega \\ \frac{-ml\cos(\theta)\sin(\theta)\omega^2 + F\cos(\theta) + (M+m)g\sin(\theta)}{l(M+m-m(\cos(\theta))^2)} \end{bmatrix}, \tag{24}$$

where horizontal force $F\,[\mathrm{N}]$ is the control input, and the values of parameters M, m, l, g are shown in Table I.

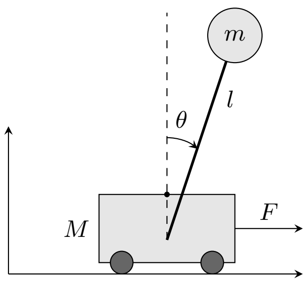


Fig. 2: Schematic illustrating the inverted pendulum on top of a cart.

The $X - Y$ position of the pendulum is given by the equations

$$c(x) := \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} p - l\sin(\theta) \\ l\cos(\theta) \end{bmatrix}. \tag{25}$$

We solve the following OCP:

$$\min_{\substack{x_0,\ldots,x_N\\ u_0,\ldots,u_{N-1}}} \ \frac{1}{2}\sum_{k=0}^{N-1} u_k^\top R_k u_k, \tag{26a}$$

$$\text{s.t.} \quad x_0 = \overline{x}_0, \tag{26b}$$

$$x_{k+1} = f(x_k, u_k), \quad k = 0, \ldots, N-1, \tag{26c}$$

$$\| [X_N - l, Y_N - l]^\top \|_2^2 - R_e^2 \leq 0, \tag{26d}$$

with $x_k = [p_k, v_k, \theta_k, \omega_k]^\top$, $u_k = F_k$, $R_k = 10^{-4}$, initial value $\overline{x}_0 = [0, 0, \pi, 0]^\top$ and X_N, Y_N the position of the pendulum at the end of the control horizon. Note that the terminal constraint is of the form $\phi(c(x_N))$, with $\phi(c) = \left\| c - \begin{bmatrix} l \\ l \end{bmatrix} \right\|_2^2 - R_e^2$, and $c(x_N)$ as in (25). We use $N = 20$ control intervals.

We compare the convergence of SCQP with that of GGN and exact Hessian SQP, in Fig. 3. We start each method close to a local solution with $R_e = 0.05\,\mathrm{m}$, plotted in Fig. (5a). Exact Hessian SQP converges quadratically, as expected. SCQP converges linearly, in contrast to GGN, for which the local solution is unstable and thus does not converge at all to the local minimum.

In fact, for decreasing radius R_e, the solution stays an unstable fixed point for GGN. This is shown in Fig 4, where we plot the largest eigenvalue of $(\widetilde{B}^\star)^{-1}(\widetilde{B}^\star - \Lambda^\star)$. Even for the terminal region quite large, as in Fig. 5b, GGN does not converge. GGN needs $R_e \approx 2\,\mathrm{m}$ for the solution to become stable; in this case, there is no swing-up. SCQP converges to a nearby local solution for all radii.

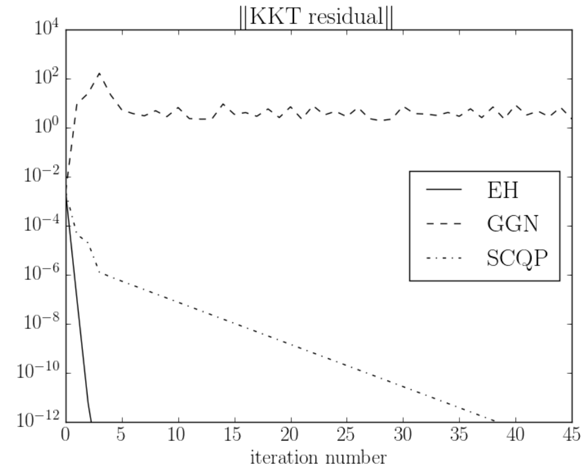


Fig. 3: Comparison of the convergence of exact Hessian SQP, GGN and SCQP to a local minimum of the pendulum OCP. The measure of convergence is the norm of the residual of the KKT system. $R_e = 0.05\,\mathrm{m}$, corresponding to "A" in Fig. 4.

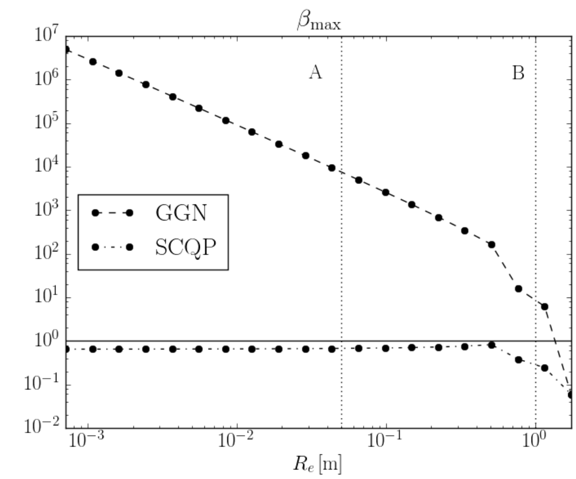
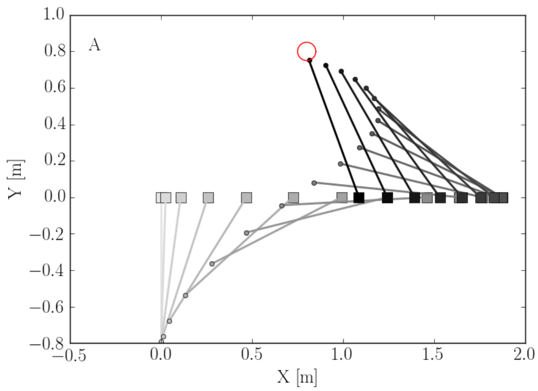


Fig. 4: Greatest eigenvalue $\beta_{\max} = \rho((\widetilde{B}^\star)^{-1}(\widetilde{B}^\star - \Lambda^\star))$ for GGN and SCQP, for different values of R_e. For a solution to OCP (26) to be a stable fixed point, $\beta_{\max}$ has to lie below the horizontal line $\beta_{\max} < 1$. The lines "A" and "B" correspond to solutions in Fig. 5.

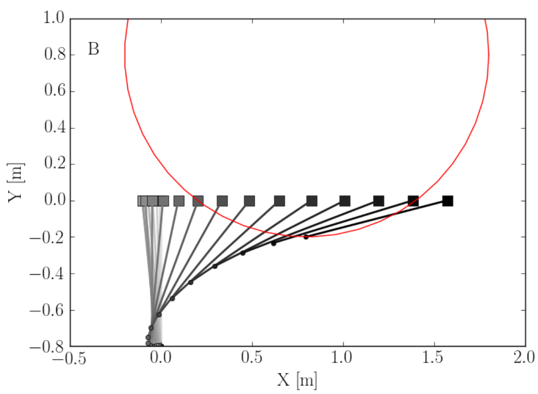
Remark 7: In this example, the extra computational cost of using SCQP instead of GGN is almost negligible. More specifically, we need: (1) $\nabla_c^2 \phi(c(x_N))$, which is constant and can be computed offline, (2) evaluation of $\frac{\partial c}{\partial x}(x_N)$, and (3) the Lagrange multiplier, which we can directly get from the QP solver.

VI. CONCLUSION

This paper proposes a generalization of the classical GGN method, referred to as the Sequential Convex Quadratic Programming (SCQP) method. Similar to Gauss-Newton, this Hessian approximation always results in a convex QP subproblem by including the second order derivatives of only convex objective and inequality constraint functions. It is shown that the SCQP approach has better local convergence properties compared to GGN. The performance of SCQP has been illustrated on a numerical case study involving a non-trivial optimal control problem.



(a) $R_e = 0.05\,\mathrm{m}$



(b) $R_e = 1.0\,\mathrm{m}$

Fig. 5: Solution of OCP (26) for different values of R_e. Note that for these radii, the solution is a stable fixed point for SCQP, and an unstable one for GGN (Fig. 4).

ACKNOWLEDGMENTS

This research was supported by the EU via ERC-HIGHWIND (259 166), FP7-ITN-TEMPO (607 957) and H2020-ITN-AWESCO (642 682), by K.U.Leuven-BOF PFV/10/002, Center-of-Excellence Optimization in Engineering (OPTEC), and the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office (DYSCO). Robin Verschueren and Niels van Duijkeren are fellows of the TEMPO Initial Training Network (FP7/2007-2013 under REA grant agreement no 607957). R. Quirynen holds a PhD fellowship of the Research Foundation – Flanders (FWO).

REFERENCES

[1] Hsl. A collection of fortran codes for large scale scientific computation. `http://www.hsl.rl.ac.uk`, 2011.

[2] CasADi. `http://casadi.org`, 2013.

[3] J. Andersson, J. Åkesson, and M. Diehl. CasADi – a symbolic package for automatic differentiation and optimal control. In S. Forth, P. Hovland, E. Phipps, J. Utke, and A. Walther, editors, *Recent Advances in Algorithmic Differentiation*, Lecture Notes in Computational Science and Engineering, pages 297–307, Berlin, 2012. Springer.

[4] H. G. Bock. Recent advances in parameter identification techniques for ODE. In *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, pages 95–121. Birkhäuser, 1983.

[5] H. G. Bock and K. J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the IFAC World Congress*, pages 242–247. Pergamon Press, 1984.

[6] H.G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, volume 183 of *Bonner Mathematische Schriften*. Universität Bonn, Bonn, 1987.

[7] M. Diehl. *Real-Time Optimization for Large Scale Nonlinear Processes*. PhD thesis, Universität Heidelberg, 2001.

[8] M. Diehl, H. G. Bock, J.P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *Journal of Process Control*, 12(4):577–585, 2002.

[9] Moritz Diehl, Andrea Walther, Hans Georg Bock, and Ekaterina Kostina. An adjoint-based SQP algorithm with quasi-Newton Jacobian updates for inequality constrained optimization. *Optimization Methods and Software*, 25(4):531–552, 2010.

[10] Hans Joachim Ferreau, Christian Kirches, Andreas Potschka, Hans Georg Bock, and Moritz Diehl. qpOASES: a parametric active-set algorithm for quadratic programming. *Mathematical Programming Computation*, 6(4):327–363, 2014.

[11] A. Griewank. *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*. Number 19 in Frontiers in Appl. Math. SIAM, Philadelphia, 2000.

[12] J. Martens and I. Sutskever. Learning recurrent neural networks with hessian-free optimization. In *28th International Conference on Machine Learning*, 2011.

[13] D. Mayne and J. Rawlings. *Model Predictive Control*. Nob Hill, 2013.

[14] Jos Luis Morales, Jorge Nocedal, and Yuchen Wu. A sequential quadratic programming algorithm with an additional equality constrained phase. *IMA Journal of Numerical Analysis*, 2011.

[15] Y. Nesterov and B.T. Polyak. Cubic regularization of Newton's method and its global performance. *Mathmatical Programming*, 112(1):177–205, 2006.

[16] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2 edition, 2006.

[17] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer-Verlag, 2000.

[18] Q. Tran-Dinh and M. Diehl. Local convergence of sequential convex programming for nonconvex optimization. In M.Diehl, F.Glineur, E. Jarlebring, and W. Michiels, editors, *Recent advances in optimization and its application in engineering*, pages 93–103. Springer-Verlag, 2010.

[19] Q. Tran-Dinh, C. Savorgnan, and M. Diehl. Adjoint-based predictor-corrector sequential convex programming for parametric nonlinear optimization. *SIAM J. Optimization*, 22(4):1258–1284, 2012.

[20] Andreas Wächter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.